

統計的検定

ネイマン - ピアソンの仮説検定ワールド から始まるモデル選択論のルーツ

三中信宏

minaka@affrc.go.jp

<http://cse.niaes.affrc.go.jp/minaka/>

〒305-8604 茨城県つくば市観音台 3-1-3

独立行政法人 農業環境技術研究所

地球環境部 生態システム研究グループ

環境統計ユニット 研究リーダー

第52回日本生態学会大会（大阪：グランキューブ大阪）

2005年3月30日（水），17:30～20:00

自由集会〈データ解析で出会う統計的問題：検定かモデル選択か〉

生態学と統計学：「震源地」の上の生態学者

「帰無モデル」論議（1970年代末～）

生態学的仮説のテスト可能性をめぐって

「モデル選択」論議（1990年代末～）

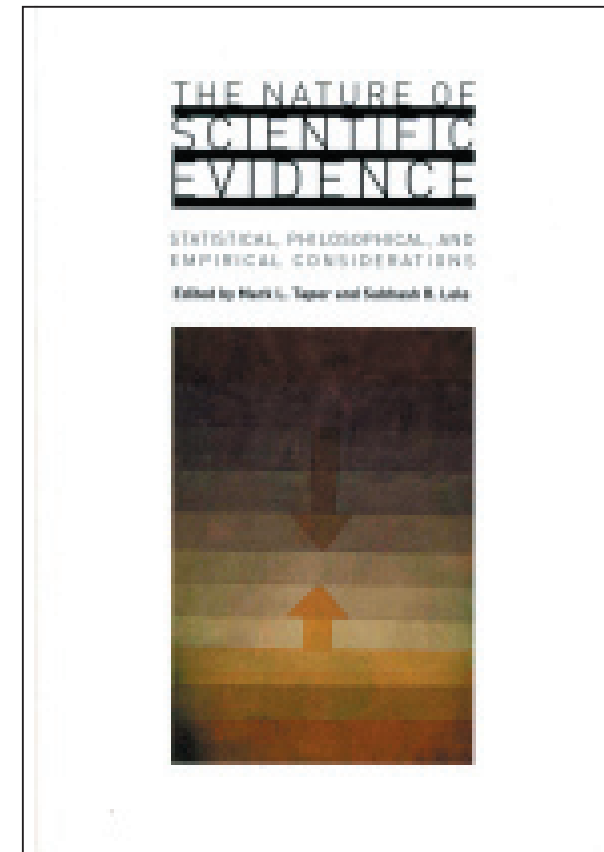
モデル選択論の観点から統計学を見直す

“蒸し返し”ではなく，“出直し”として。

Mark L. Taper & Subhash R. Lele (eds.)

*The Nature of Scientific Evidence: Statistical,
Philosophical, and Empirical Considerations*

(2004, The University of Chicago Press)



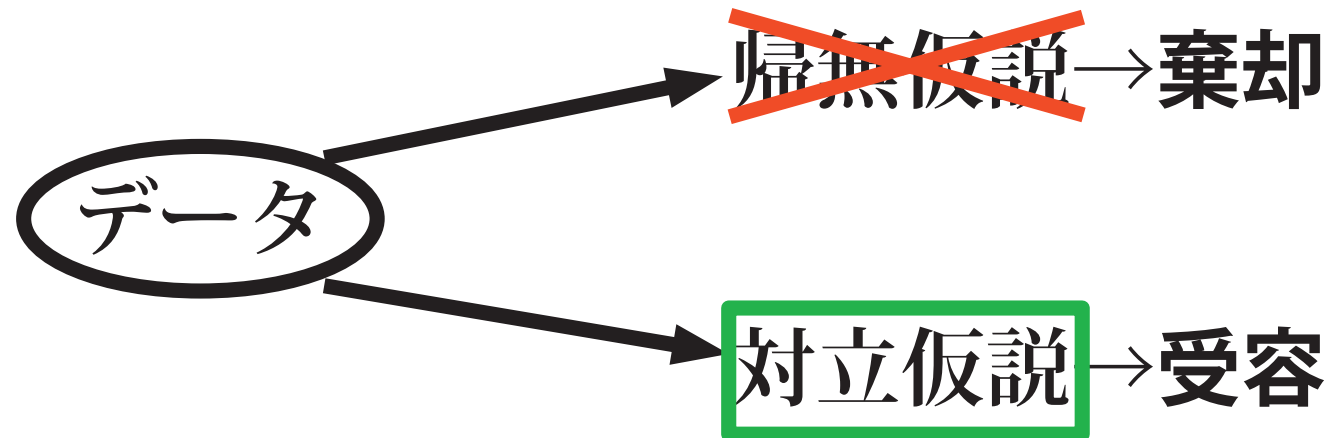
第 52 回日本生態学会大会 (大阪：グランキューブ大阪)

2005 年 3 月 30 日 (水), 17:30 ~ 20:00

自由集会 〈データ解析で出会う統計の問題：検定かモデル選択か〉

「現在，科学において行なわれている統計分析の大半は，もともと意思決定（decision-making）の道具として開発された方法—たとえば，Fisherの有意义検定やNeyman-Pearsonの仮説検定—を用いている．科学の中でこれらの手法はどの仮説を選ぶかの決定を下すために使われている」（p. xv）

「意思決定パラダイム」の中の統計学



Data as Executioner

仮説検定史

- **Pure Significance Test** (R. A. Fisher 1925)

ある仮説 H のデータ x に対する「整合性」の程度を調べる。
有意確率： $p = \Pr[t(X) > t(x)/H]$ が小さいほど整合性が低い。
仮説の棄却・受容に関する意思決定ではない。

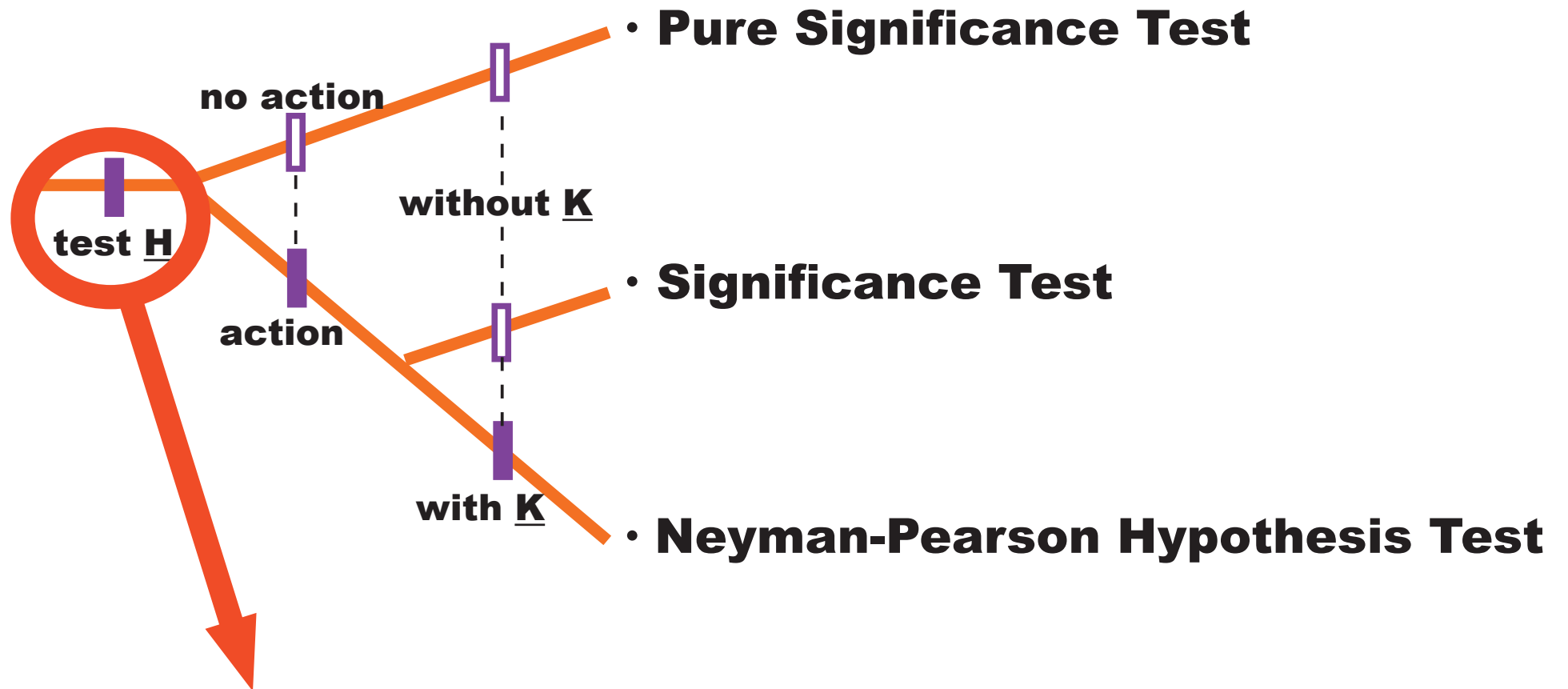
- **Significance Test**

ある「有意水準」 α のもとで、仮説 H の棄却・受容を意思決定する。
※ 「 $p < \alpha$ のとき、 H を棄却する」という行動基準を置く。

- **Neyman-Pearson Hypothesis Test**

(J. Neyman & E. S. Pearson 1933)

ある帰無仮説 H に対して、対立仮説 K を設定し、有意水準 α のもとでの仮説を棄却・受容するための検定を構築する。尤度比の導入。検定力 (power) の重視などなど。



仮説検定クレードの共有派生形質

Neyman-Pearson 的に「仮説をテストする」ためには：

- 1) モデルを指定する.
- 2) そのモデルに含まれるパラメータ θ に関する帰無仮説 H と対立仮説 K を設定する.
- 3) 有意水準 (type I error) α を指定する.
- 4) 検定の手続きをつくる. 尤度比と検定力に着目.
- 5) 以上のすべてをデータ収集に先立って行なう.

われわれは
Neyman-Pearson 仮説検定を
必ずしも “正しく” 使っていない
のではないか？

一般化線形モデル
系統樹推定
分子進化プロセス

.....

テスト対象である仮説の
占めるべき「地位」が
変わりつつあるのではないか？

絶対ランキング→相対ランキング

(true or false)

(better or worse)



Rudolf Carnap [帰納論理学]

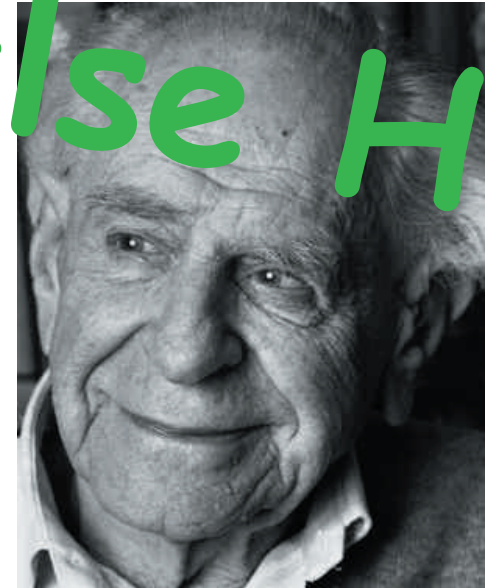
データによる仮説の確証
(confirmation) の方法をつくろうとした。

H

False H

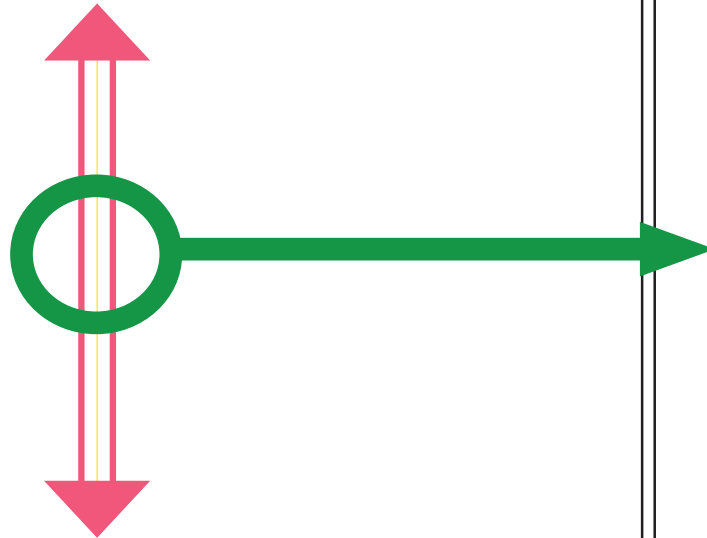
Karl R. Popper [仮説演繹主義]

データによる仮説の反証
(falsification) の方法をつくろうとした。



帰納主義 (Carnap)

→真なる仮説の受容



反証主義 (Popper)

→偽なる仮説の棄却

仮説発見的推論

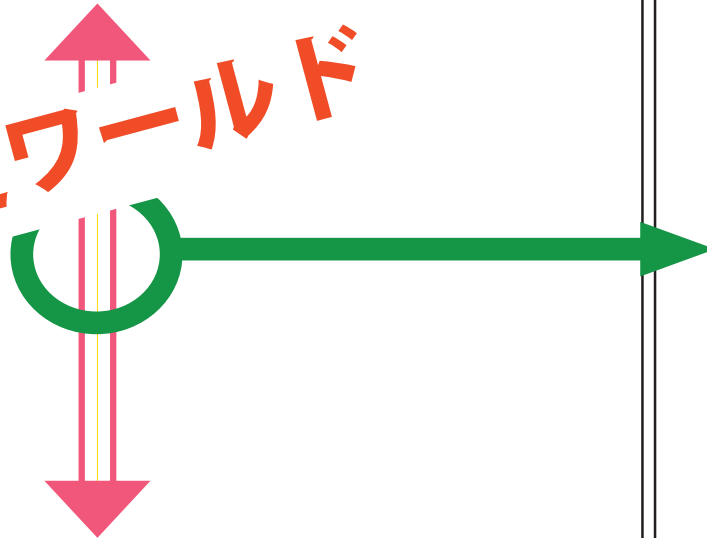
Abductive inference

→その時点でベスト
の仮説の探索

帰納主義 (Carnap)

→真なる仮説の受容

仮説検定ワールド



反証主義 (Popper)

→偽なる仮説の棄却

仮説発見的推論

Abductive inference

→その時点でベスト

の仮説のベスト

モデル選択ワールド

Elliott Sober

観察されたデータから，ある仮説が真であること（強確証），または偽であること（強反証）を結論するのは現実的ではない．むしろ，データがそれぞれの対立仮説をどれくらい支持するか（弱確証／弱反証）を比較する必要がある．

『過去を復元する』1996, 蒼樹書房]





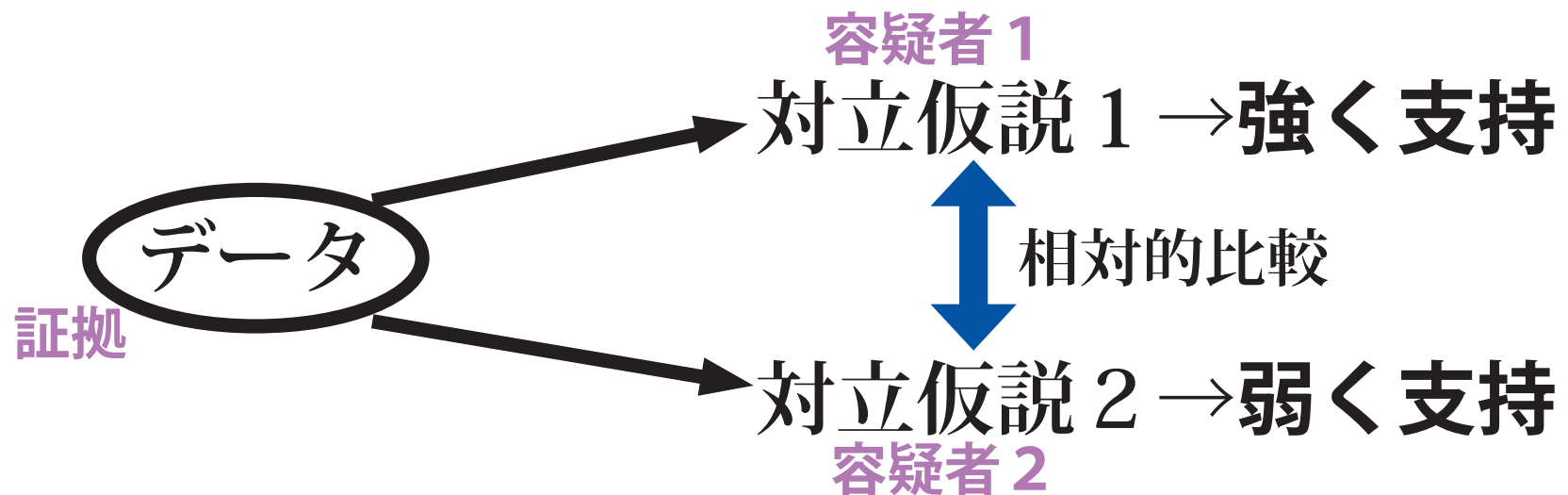
Richard Royall

どのような状況で，与えられた観察が，ある統計的仮説ではなく，別の仮説を支持すると言えるのか．この根本問題を統計学は長い間無視し続けてきた．それは，Neyman-Pearson 統計学が，**データを仮説の証拠とみなす**のではなく，意思決定の理論として定式化されてきたからである．

[『Statistical Evidence: A Likelihood Paradigm』 1997, Chapman & Hall]

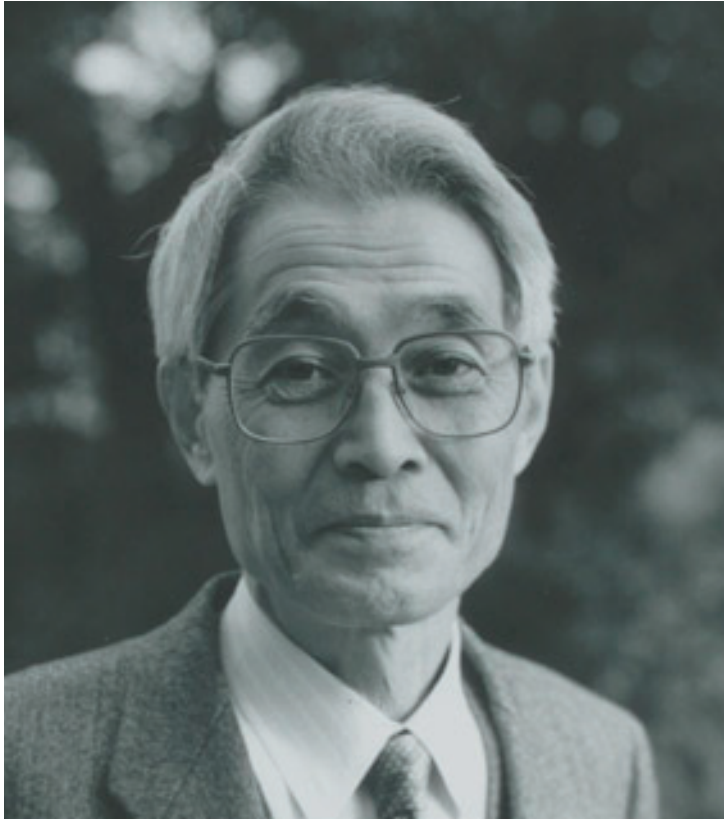
「法科学的パラダイム」の中の統計学

forensic



Data as Evidence





良いモデルは
「予測確度」
(predictive accuracy)
が大きい

[赤池弘次]

第52回日本生態学会大会（大阪：グランキューブ大阪）

2005年3月30日（水），17:30～20:00

自由集会〈データ解析で出会う統計的問題：検定かモデル選択か〉

現場の科学者は、すでに「モデル選択」的な考え方をすでに受容しつつあるのではないか。その際、Neyman-Pearson 流の「仮説検定」のスタイルを“やりくり流用”して、出てきた結果を「モデル選択」的に解釈するということもあり得るだろう。そういうユーザーには、「あなたのやっていることは、仮説検定ではなく、実はモデル選択なのですよ」と言ってあげると、きっと憑きものは落ちるにちがいない。

粕谷さん、あとはよろしく！